DOCUMENT RESUME

ED 303 722 CG 021 420

AUTHOR Schmidt, Frank L.; And Others
TITLE General Cognitive Ability vs. General and Specific Aptitudes in the Prediction of Training Performance: Some Preliminary Findings.
PUB DATE 15 Aug 88
NOTE 18p.; Paper presented at the Annual Meeting of the American Psychological Association (96th, Atlanta, GA, August 12-16, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Ability; *Aptitude Tests; *Cognitive Ability; Military Personnel; *Military Training; *Performance; *Predictive Validity
IDENTIFIERS *Armed Services Vocational Aptitude Battery

ABSTRACT

Recently there appears to have been an increase in interest in the relative power of general ability and narrower cognitive aptitudes to predict real world performance in training programs and on the job. This area has important practical implications for personnel selection and classification, particularly for large organizations such as the United States military which assigns people differentially to jobs based on patterns of measured abilities. This study examined the ability of Armed Services Vocational Aptitude Battery (ASVAB), forms 8, 9, and 10 to predict training performance, using 10 technical jobs in the U.S. Navy with at least 900 people in each job. Performance was the measured criterion. For one job, Mess Management Specialist, performance was measured as final school grade. The other nine technical jobs were self-paced; the criterion was the number of hours required to complete training. General mental ability alone did about as well as differential weighting at the level of the three general aptitudes (quantitative, verbal, technical). Using all 10 specific aptitude tests as separate predictors increased validity by about eight percent. Future analysis will examine the specific values of beta weights and will be testing Hunter's (1983) path models for fit to the data from individual jobs. (ABL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED303722

General Cognitive Ability vs. General and Specific Aptitudes in The Prediction of Training Performance: Some Preliminary Findings

Frank L. Schmidt
University of Iowa

John E. Hunter
Michigan State University

Mark Larson
University of Iowa

ED301420

Presented as part of the Symposium "Promising Areas for Psychometric Research". W.H. Augoff, Chair. 1988 American Psychological Association Convention, Atlantic, GA, August 15, 12:00-1:50 p.m., Atlanta Marriott Marquis, International Salon F.

14mlscm/7/22/88

General Cognitive Ability Vs. General and Specific Aptitudes in The Prediction of Training Performance: Some Preliminary Findings

Recently there appears to have been an increase in interest in the relative power of general mental ability and narrower cognitive aptitudes to predict real world performance in training programs and on the job (e.g., see Thorndike, 1986; Hunter, 1986). This question has important implications for theories of human cognitive abilities. If narrower abilities add nothing to prediction over general ability, then the status of narrower abilities within theories of ability will have to be reconsidered. In addition, it has important practical implications for personnel selection and classification, particularly for large organizations like the U.S. military which assign people differentially to jobs based on patterns of measured abilities.

Data relevant to the question of the relative predictive power of general mental ability and narrower abilities comes from the civilian, as well as the military sector (Hunter, 1980; Jensen, 1984, 1987). This paper, however, is concerned with research on military personnel. Recent research by Hunter (1983; 1984; 1985) based on very large military samples appears to indicate that general cognitive ability is as good or better a predictor of performance in training in most military job families as ability composites derived specifically to predict success in particular job families. These findings are contrary to the current theory that is the foundation of differential assignment of personnel to jobs in the military. That theory, differential aptitude theory (or specific aptitude theory), postulates that specific aptitude factors assessed by particular

tests or by clusters of tests make an incremental contribution to the prediction of performance over and above the contribution of general cognitive ability.

Cast in causal terms, differential aptitude theory postulates that specific aptitudes have a causal impact on performance over and above any causal impact of general cognitive ability. Hunter (1983) showed that when average validities across jobs for each test are used, the causal model implied by differential aptitude theory does not fit large sample military data sets. Instead, the model that fits the data best is one in which the only ability causing performance is general cognitive ability and in which aptitudes are themselves caused by general cognitive ability. No causal paths to performance from either general aptitudes (Verbal, Quantitative, and Technical; defined by clusters of tests) or specific aptitudes (defined by single tests) were necessary to fit the data. This theory of the underlying processes causing performance predicts that for military job families, general cognitive ability would predict performance at least as well as regression-based composites of specific aptitude derived to predict performance in the particular job family. In his analyses, this prediction was generally borne out.

Implications for the Utility of Classification Systems in the Military

These findings have profound implications for the utility of the complex systems used in the military for the differential assignment of personnel to jobs (personnel classification systems). In order for a classification system to have utility beyond that created by the selection

aspect of the system (i.e., beyond the utility resulting from rejection of generally unpromising applicants), the regression equations for predicting performance must be different for different jobs or job families. In traditional military classification systems, both test scores and job performance have been standardized, and mean levels of job performance have been implicitly assumed to be equal across jobs (Brogden, 1959). Given these assumptions and Hunter's findings, if the validity of general cognitive ability is equal for all jobs or job families, then all (standard score) regression equations are identical, to wit:

$$\hat{z}_{y_i} = rz_{x_i}$$

where: $\hat{z}_{y_i}$ = predicted performance in standard score form for applicant i.

r = the validity of general cognitive ability; and

z_{x_i} = the standard score on general cognitive ability for applicant i.

Under these circumstances, personnel classification has no value over and above that of selection. The actual findings were not quite this drastic, because the validity of general mental ability was found to vary somewhat across job families. For example, the validity was found to somewhat lower for the Combat job family and for the Equipment Operator job family (Hunter, 1983). This variation in validity means that the standard score regression equations are not identical and that there is some small room left for differential assignment to have utility. (Differential assignment

based on a single predictor is called placement; Cronbach and Gleser, 1965.)

Nevertheless, Hunter's findings imply that the utility of military classification systems is at best only a small fraction of what it has been assumed to be. Because this implication is so profound, it is advisable to conduct further tests of his findings.

Hunter's path analytic studies were conducted using average validities across all jobs for which validity data were available; these studies led to the prediction that general cognitive ability should have higher validity than regression-based composites of specific aptitudes for every job. However, individual jobs with sample sizes large enough to test this prediction were not available. Because of this, Hunter (1983) used existing military job families as his unit of analysis, with each job family being represented by the average validities of the composites (and of general cognitive ability) across the jobs making up the job family. If the job families had been developed by a means that ensured that the jobs in each family would be homogeneous for aptitude requirements (if there were indeed such differential requirements) this approach would have had no problems. However the job families were originally created by a method that capitalizes heavily on chance configurations in the beta weights, i.e., that capitalizes on sampling error in the beta weights (Hunter 1983). Using samples that were often small in relation to the number of tests, jobs that had similar patterns of beta weights were grouped together into job families. To the extent that job families were based on random sampling error in the beta weights, the jobs in any given job family will

be a random sample of all jobs with respect to aptitude requirements (if there are differential aptitude requirements). Although complete randomness is almost certainly not the case, there is undoubtedly a substantial random component (as noted by Hunter, 1983), and this would serve to blur any underlying differential aptitude requirements.

This problem can be solved by testing the differential aptitude theory using individual jobs that have very large samples. If Hunter's findings are verified using such a data base, his conclusions would appear to be unshakable and should then be incorporated into military human resources procedures and policies. This would probably entail: (1) seeking a new basis for classification utility in differences among jobs in average value to the military and in within-job variability of performance value to the military; (2) seeking new predictors that might create differential validity (e.g., psychomotor tests); and (3) a focus on increasing and measuring the utility of the selection component alone (which is substantial). On the other hand, if differential aptitude theory is supported, an obvious need will be the development of better methods of creating job families in order to take advantage of classification utility resulting from differential prediction. This follows from two facts which prevent the use of differential prediction at the level of individual jobs: (1) for many jobs, sample sizes will probably never be large enough for reliable determination of differential weights; and (2) even if the necessary sample sizes were available, for practical reasons operational classification systems must be based on a smaller number of job groupings (i.e., job families).

14mlscm/7/22/88

There are three levels of ability generality - specific aptitudes (assessed by individual tests), general aptitudes (assessed by clusters of tests identified by confirmatory factor analyses; e.g., verbal ability), and general cognitive ability (assessed by a cluster of general aptitudes identified by confirmatory factor analysis). Traditional operationalization of differential aptitude theory in the military has focused on specific aptitudes. Regression-based composites of scores on individual tests have been derived to predict performance in individual jobs and job families. Assuming the validity of differential aptitude theory, this approach would be satisfactory only if every test measured a different aptitude. If this were not the case, and if, for example, the Work Knowledge Test and the Paragraph Comprehension Test of the ASVAB both measured the same aptitude (verbal ability), then the beta weight for verbal aptitude would be split between the two tests in the regression analysis, with the specific split depending on sampling error in the validities and test intercorrelations (assuming equal test reliabilities). Since intercorrelations among tests measuring the same aptitude would be high, computed beta weights would have large standard errors, and would be unstable from sample to sample. Stability aside, they would also be difficult to interpret correctly (e.g., in our example, in terms of the importance of verbal aptitude). That is, if differential aptitude theory were valid at the level of general aptitudes but not valid at the level of specific aptitudes, then beta weights for individual tests would be both unstable and difficult to interpret.

Under these circumstances, factors specific to general aptitudes, but not those specific to individual tests, would contribute to validity over and above the contribution of general cognitive ability. In such a case, the problems described above could be avoided by conducting the analyses at the level of general aptitudes, rather than at the level of individual tests. The expected finding would then be that regression-weighted composites of general aptitudes would predict performance better than general cognitive ability alone. Because of these considerations, differential aptitude theory should be tested at the level of general aptitudes as well as at the level of individual tests.

Method

The tests used were the 10 subtests of The Armed Services Vocational Aptitude Battery (ASVAB), forms 8, 9 and 10. The ASVAB was administered prior to entry into the Navy; all validities are predictive. These tests are described in Table 1; reliabilities are also given.

The preliminary study reported here is based on 10 technical jobs in the U.S. Navy. The goal was to have data for 1,000 or more people for each job; however, two jobs were included with Ns of less than 1,000 (Ns of 958 and 928). The jobs and their sample sizes are shown in Table 3.

The criterion was performance in Navy technical schools. For one job (Mess Management Specialist), performance was measured in the traditional manner: as final school grade (FSG). FSG is based on the average achievement tests administered during training. The other nine technical schools were self-paced; the criterion was the total number of hours required to complete training.

Because interest is in the predictive power of the tests in the applicant group, all validities were corrected for multivariate range restriction by the Navy. Intercorrelations among the tests were estimated on a sample of over 140,000 Navy applicants.

The predictive power of specific aptitudes was estimated by multiple regression using all 10 tests.

Hunter (1983; see also Kass et al., 1982) found that there are three general cognitive aptitudes underlying The ASVAB: Quantitative, Verbal and Technical. The tests that best measure these general aptitudes are:

$$Q = AR + MK$$

$$V = WK + GS$$

$$T = MC + EI$$

In this study, Q, V and T were measured by these standard score (equally weighted) sums. The predictive power of general aptitude theory was assessed by regressing criterion scores on Q, V and T composites. General mental ability is measured by the sum of the three standardized cognitive aptitude composites:

$$G = Q + V + T, \quad \text{or}$$

$$G = (AR + MK) + (WK + GS) + (MC + EI)$$

These relations are summarized in Table 2.

The multiple Rs for the 10 tests and for the three general aptitude composites were shrunken using the Wherry formula (Catlin, 1980) to adjust for capitalization on chance. However, the sample sizes used in the shrinkage formula had to be adjusted to take into account the fact that the multivariate range restriction correction increases the standard error of

the validity coefficients, which reduces the effective sample size. For each job, the sampling error variance of the average validity was estimated as $(\bar{r}_c/\bar{r})^2 s_e^2 - s_{e_c}^2$, where:

$\bar{r}_c$ - the mean range corrected validity

$\bar{r}$ - the mean uncorrected validity

s_e^2 - the mean sampling error variance prior to the range correction.

$s_{e_c}^2$ - the mean sampling error variance of the range corrected validities.

The effective sample size (N_E) was then computed for each job as follows:

$$N_E = \frac{(1 - \bar{r}^2)^2 + s_{e_c}^2}{s_{e_c}^2}$$

This formula is obtained by solving the standard sampling error variance formula for r for N. The adjusted Ns were considerably smaller than the nominal Ns, as shown in Table 3. Sampling error variance is increased considerably by range restriction corrections. This adjustment to N is only an approximation, but no more exact method appears to be known at the present time.

It should be noted that the Wherry formula estimates the population correlation that would be produced by using the population weights themselves, i.e., the exact values of those weights and not estimates of them. In real testing programs, the exact population weights (which are optimal) are never known. Only estimates are available, and since estimate

depart from the true values, they produce a population correlation that is smaller. Thus our results for the 10 tests and QVT are upper bound values for validities; in practice, smaller values would be expected. This is not true for G, because the correlation for G is a zero order correlation.

Results and Discussion

The results are best summed up by the means at the bottom of Table 3. On the average, differentially weighting the V, Q and T composites using population regression weights is estimated to increase validity only by .01 or 2%. This tiny increase may not even be replicable. Thus, on average, general mental ability alone does about as well as differential weighting at the level of the three general aptitudes. This finding might be surprising to those who would expect that prediction could be enhanced for these technical jobs by assigning larger weights to Technical Ability (and perhaps also to Quantitative Ability).

Using all 10 tests as separate predictors, as suggested by specific aptitude theory, increases validity by an average of .04 or 8%. While this increment may not seem impressive, it is four times larger than the average QVT increment and, if real, would be of practical value. However, it is by no means certain at this point that the increment is real. This study is only preliminary, and we will be looking at more (and better) data in the future. Data to be analyzed in the future is based on traditional training school grades rather than time taken in training. Previous research indicates relations between test scores and traditional grades are linear or close to linear. This may not be true for the time taken in training measure, and so far it has not been possible to test for linearity in those

data. If there are violations of linearity, these, in combination with multivariate range restriction corrections, could distort findings. So more information is required before concluding that specific aptitude theory is supported over general aptitude theory and general ability theory.

The above results are for observed scores. The more theoretically minded among you may wonder how the findings might look at the true score level. That is, if we were hypothetically able to measure each of the 10 specific aptitudes with perfect reliability, and measure the Q, V, T and G factors perfectly reliably, how would the results look? This analysis was conducted, correcting for unreliability in each of the 10 tests using the figures in Table 1. Confirmatory factor analyses were used to estimate the intercorrelation of factors and true score validity for Q, V, T and G. As expected, all correlations increased. However, the relative performance of specific aptitude theory, general aptitude theory, and general ability theory was very much the same as shown in Table 3 for observed scores.

In future analysis, we will also be examining the specific values of beta weights and will be testing Hunter's (1983) path models for fit to the data from individual jobs. Analysis will be run for both observed and true scores.

Because the data requirements for strong tests of these three theories are very stringent, and because there are some complex and tricky statistical and measurement problems involved in these tests, we do not yet have final answers. But we feel that progress has been made and will continue to be made.

14mlscm/7/22/88

Table 1
Predictor Variables From ASVAB Forms 8, 9, and 10

Test[a]	Abbreviation	Reliability[b]	Description
General Science	GS	.86	A 25-item test of knowledge of the physical (13 items) and biological (12 items) sciences--11 minutes.
Arithmetic Reasoning	AR	.91	A 30-item test of ability to solve arithmetic word problems--36 minutes.
Word Knowledge	WK	.92	A 35-item test of knowledge of vocabulary, using words embedded in sentences (11 items) and synonyms (24 items)--11 minutes.
Paragraph Comprehension	PC	.81	A 15-item test of reading comprehension --13 minutes.
Numerical Operations	NO	.78	A 50-item speeded test of ability to add, subtract, multiply, and divide one- and two-digit numbers--3 minutes.
Coding Speed	CO	.85	An 84-item speeded test of ability to recognize numbers associated with words from a table--7 minutes.
Auto and Shop Information	AS	.87	A 25-item test of knowledge of automobiles, shop practices, and use of tools--11 minutes.
Mathematics Knowledge	MK	.87	A 25-item test of knowledge of algebra, geometry, fractions, decimals, and exponents--24 minutes.
Mechanical Comprehension	MC	.85	A 25-item test of knowledge of mechanical and physical principles--19 minutes.
Electronics Information	EI	.82	A 20-item test of knowledge of electronics, radio, and electrical principles and information--9 minutes.

[a] Reported as Navy Standard Scores having a mean of about 50 and a standard deviation of 10 for an unrestricted recruit population.

[b] KR-20 reliabilities for power tests; parallel form reliabilities for No and CS; from Kass et al. (1982).

14mlscm/7/22/88

Table 2
Definitions of General Aptitudes and
General Mental Ability

General Aptitudes

Quantitative:

$$Q = AR + MK$$

Verbal (Conceptual):

$$V = WK + GS$$

Technical:

$$T = MC + EI$$

General Mental Ability:

$$G = Q + V + T$$

$$G + (AR + MK) + (WK + GS) + (MC + EQ)$$

Note: See Table 1 for test names.

Table 3
Estimated Validities Using 10 Tests, 3 Aptitudes
and General Mental Ability Alone

	Job (Navy Titles)	Sample Size Nominal	Sample Size Adj.[a]	Validities[b] 10 Tests	Validities[b] QVT	Validities[b] G
1.	Mess Mgmt. Spec. (MS)	1581	909	.59	.58	.56
2.	Aviation Elec. Tech. (AT)[c]	1489	338	.46	.42	.43
3.	Aviation Elec. Tech. (AT)[c]	2245	679	.78	.75	.75
4.	Electronics Tech. (ET)[d]	958	112	.70	.66	.64
5.	Electronics Tech. (ET)[d]	928	207	.61	.61	.59
6.	Boiler Tech. (BT)	2085	1353	.48	.44	.44
7.	Engineman (EN)	1258	583	.44	.39	.38
8.	Machinists Mate (MM)	2598	1181	.37	.36	.35
9.	Aviation Elec. Mate (AE)	1606	651	.54	.53	.53
10.	Electricians Mate (EM)	1109	158	.54	.53	.53
			Means	.56	.53	.52
			Percent Increase over G	8%	2%	--

[a]Adjusted to the effective N allowing for the increase in the standard error of the validities resulting from the corrections for multivariate range restriction.

[b]Multiple Rs for the 10 tests and QVT are shrunken using the Wherry formula.

[c,d]Independent samples from the same job; the explanation for the disparities is unknown.

14mlscm/7/22/88

References

Brogden, H. E. (1959) Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. Educational and Psychological Measurement, 19, 181-190.

Booth-Kewley, S., Foley, P. & Swanson, L. (1984a) Predictive Validation of the Armed Services Vocational Aptitude Battery (ASVAB) form 8, 9, and 10 Against Performance in 100 Navy Schools. Technical Report 85-15, Navy Personnel Research & Development Center, San Diego, CA 92152-6800.

Booth-Kewley, S., Foley, P. & Swanson, L. (1984b) Predictive Validation of the Armed Services Vocational Aptitude Battery (ASVAB) form 8, 9, and 10 Against Performance in 100 Navy Schools. Technical Supplement to TR85-15, Navy Personnel Research & Development Center, San Diego, CA 92152-6800.

Catlin, P. (1980) Estimation of the predictive power of a regression model. Journal of Applied Psychology, 654, 407-414.

Cronbach, L. J. & Gleser, G. C. (1965) Psychological tests and personnel decisions. Second Edition. Urbana, Illinois: University of Illinois Press, 1965.

Hunter, J. E. (1980) Validity generalization for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitudes Test Battery (GATB). Washington, D.C.: U.S. Employment Service, Department of Labor.

Hunter, J. E. (1983) The prediction of job performance in the military using ability composites: The dominance of general cognitive ability over specific aptitudes. Report for Research Applications, Inc., in partial fulfillment of DOD contract No. F41689-83-C-0025.

Hunter, J. E. (1984) The validity of the Armed Forces Vocational Aptitude Battery (ASVAB) High School Composites. Report for Research Applications, Inc., in partial fulfillment of DOD contract No. F41689-83-C-0025.

Hunter, J. E. (1985) Differential validity across jobs in the military. Report for Research Applications, Inc., in partial fulfillment of DOD contract No. F41689-83-C-0025.

14mlscm/7/22/88

Hunter, J. E. (1986) Cognitive ability, cognitive aptitudes, job knowledge, and job performance. Journal of Vocational Behavior, 29, 340-362.

Hunter, J. E. & Schmidt, F. L. (1982) Fitting people to jobs: Implications of personnel selection for national productivity. In E. A. Fleishman and M. D. Dunnette (Eds.) Human Performance and Productivity. Volume I: Human Capability Assessment. Hillsdale, N.J.: Earlbaum, 233-284.

Jensen, A. R. (1984) Test validity: g versus the specificity doctrine. Journal of social and behavioral structures, 7, 93-118.

Jensen, A. R. (1987) The g beyond factor analysis. In R. R. Ronning and J. C. Witt (Eds.), The influence of cognitive psychology on testing. Hillsdale, N.J.: Erlbaum.

Kass, R. A., Mitchell, K., Grafton, F. and Wing, H. (1982) Factor structure of the ASVAB, Forms 8, 9, and 10: 1981 army applicant sample. U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

Schmidt, F. L., Hunter, J. E. (1983) Individual differences in productivity: An empirical test of estimates derived from studies of selection procedure utility. Journal of Applied Psychology, 68, 407-415.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979) The impact of a valid selection procedure on work-force productivity. Journal of Applied Psychology, 64, 609-626.

Thorndike, R. L. (1986) The central role of general ability in prediction. Multivariate Behavioral Research, 20, 241-254.

Thorndike, R. L. (1986) The role of general ability in prediction. Journal of Vocational Behavior, 29, 332-339.

U.S. Department of Navy (1966) U.S. Navy occupational handbook. Washington, D.C.: Author.

U.S. Navy (Sept. 1975) Manual of Navy enlisted manpower and personnel classification and occupational standards. Washington, D.C.: Author.

14mlscm/7/22/88